High Level Design (LLD)

Concrete Compressive Strength Prediction (Machine Learning)

By

Kavitha Narsapur

Abstract

The quality of concrete is determined by its compressive strength, which is measured using a conventional crushing test on a concrete cylinder. The strength of the concrete is also a vital aspect in achieving the requisite longevity. It will take 28 days to test strength, which is a long period. Thus this project aims to predict concrete strength using machine learning models. Different models like Linear regression, Lasso regression, Decision tree regression, Random forest regression, Ada boost regression, Gradient boost regression, XGBoost regression and Voting regression were built. Out of all the regressors, XGBoost regression resulted in highest R2_score of 92.4635% and low RMSE of 4.787508. Hence XGBoost regression can be used for predicting concrete compressive strength.

Introduction

Why this High-Level Design Document?

This High-level Design (HLD) Document indicates all the necessary steps that were carried out prior to building machine learning model as Data pre-processing and Exploratory data analysis. After data preparation, the document indicates which models were built and tested on test data and also describes model deployment.

Scope

The HLD documentation presents the structure of the system, such as the application architecture (layers), application flow (Navigation), and technology architecture. The LLD uses non-technical to mildly-technical terms which should be understandable to the administrators of the system. This software system will be a Web application. This system will be designed to predict concrete compressive strength.

General Description

Introduction

Thus this project aims to predict concrete strength using machine learning models.

Problem Statement

The quality of concrete is determined by its compressive strength, which is measured using a conventional crushing test on a concrete cylinder. The strength of the concrete is also a vital aspect in achieving the requisite longevity. It will take 28 days to test strength, which is a long period. Thus this project aims to predict concrete strength using machine learning models.

Approach

This project involves data pre-processing like Data cleaning, Exploratory data analysis, Model building and Model deployment. Different models like Linear regression, Lasso regression, Decision tree regression, Random forest regression, Ada boost regression, Gradient boost regression, XGBoost regression and Voting regression were built and then tested on to test data.

Dataset overview

Number of instances: 1030

Number of Attributes: 9

Attribute breakdown 8 quantitative input variables, and 1 quantitative output variable

Missing Attribute Values: None

Given are the variable name, variable type, the measurement unit and a brief description. The concrete compressive strength is the regression problem. The order of this listing corresponds to the order of numerals along the rows of the database.

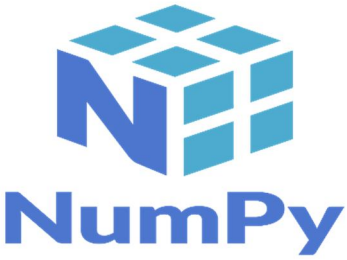
Name -- Data Type -- Measurement -- Description

Cement (component 1) -- quantitative -- kg in a m3 mixture -- Input Variable
Blast Furnace Slag (component 2) -- quantitative -- kg in a m3 mixture -- Input Variable
Fly Ash (component 3) -- quantitative -- kg in a m3 mixture -- Input Variable
Water (component 4) -- quantitative -- kg in a m3 mixture -- Input Variable
Superplasticizer (component 5) -- quantitative -- kg in a m3 mixture -- Input Variable
Coarse Aggregate (component 6) -- quantitative -- kg in a m3 mixture -- Input Variable
Fine Aggregate (component 7) -- quantitative -- kg in a m3 mixture -- Input Variable
Age -- quantitative -- Day (1~365) -- Input Variable
Concrete compressive strength -- quantitative -- MPa -- Output Variable

Further Improvements

The performance of machine learning models can be further enhanced by tuning various hyperparameters of the models.

Tools used


NumPy


pandas


matplotlib


seaborn

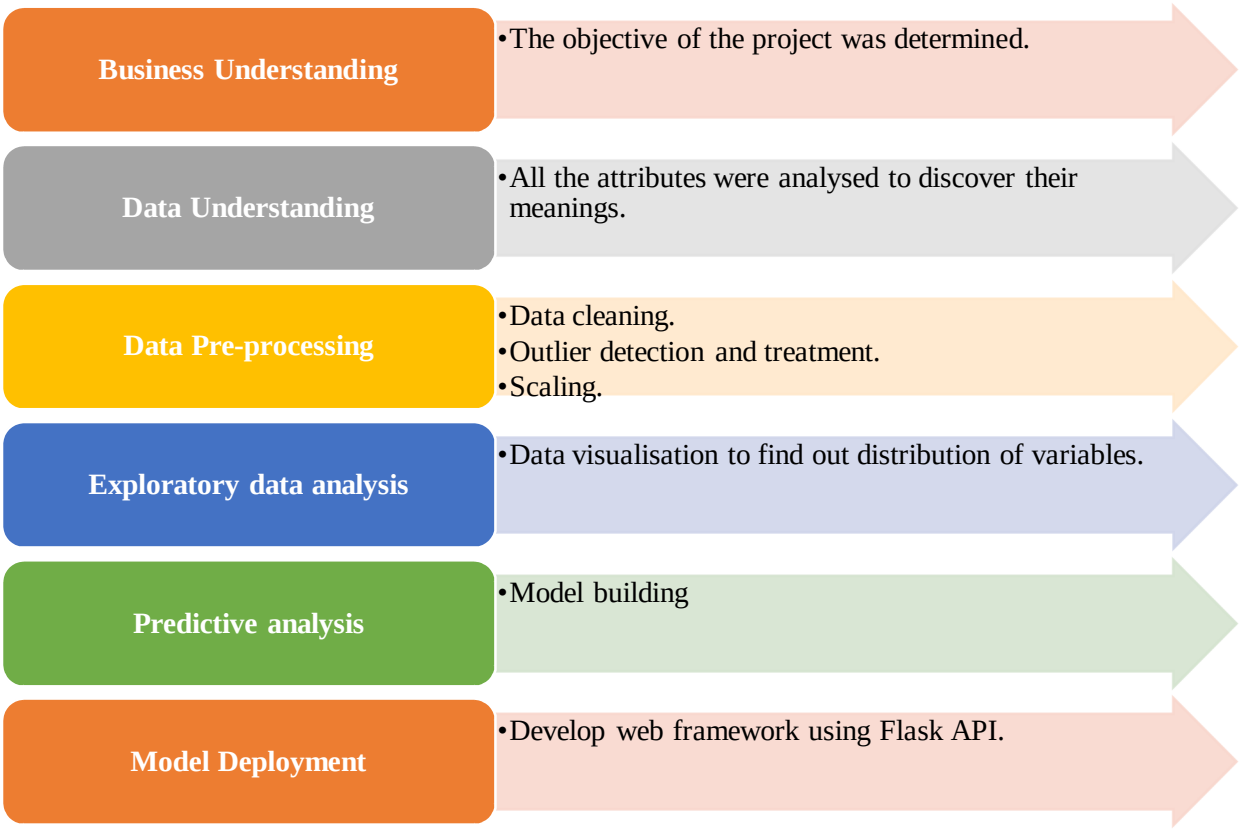

scikit
learn


Pickle


Flask


python

Architecture



Conclusion

Out of all the tested models, XGBoost regression resulted in highest R2_score of 92.4635% and low RMSE of 4.787508. Hence XGBoost regression can be used for predicting concrete compressive strength.